A logo of a university

Description automatically generated

**DSCI-6011-02 DEEP LEARNING**

**Final Project ON**

**Precision Segmentation of Retinal Nerve Fibers in Fundus Images using Deep Learning**

**Done BY**

**REVANTH PAINENI(**[**rpain1@unh.newhaven.edu**](mailto:rpain1@unh.newhaven.edu)**)**

**SRIDHAR YEMUNDLA(**[**syemu1@unh.newhaven.edu**](mailto:syemu1@unh.newhaven.edu)**)**

**Contents**

**[Abstract:](#_Toc7043)** [3](#_Toc7043)

[**Introduction:** 3](#_Toc7044)

[**Related Work:** 4](#_Toc7045)

**Methods**[**:** 5](#_Toc7046)

[Data Preprocessing: 5](#_Toc7047)

[Model Architectures: 7](#_Toc7048)

[**Experimental Setup:** 9](#_Toc7049)

[**Metrics for Evaluation:** 9](#_Toc7050)

[**Results:** 9](#_Toc7051)

[**Discussion:** 14](#_Toc7052)

[**Conclusion:** 16](#_Toc7053)

**References**:………………………………………………………………………………………………..17

**Abstract:**

This project demonstrates how to use different types of architectures on a retinal image segmentation annotation project and evaluates them through four types of metrics monitor. First, we utilize two types of pretrained models such as EfficientNet and MobileNet that have remained popular in computer vision tasks for past years. Second, we utilize a hybrid model that combines both good feature extraction capabilities via U-Net backbone and information sharing ability through attention mechanisms. For these models, we will monitor four types of metrics such as accuracy (counts ratio of correctly segmented pixels), binary cross entropy loss. Finally, for the robustness of our model, we have used two different datasets DRIVE and HRF.

**Introduction:**

It is important to precisely segment the retinal images, but it is very difficult as classifying the internal structure of the nerve segment is complex, mostly vanilla architectures not able to handle these complexities. We are using pretrained architectures like EfficientNet, MobileNet, etc. We are using Hybrid U-Net with attention mechanism.

We’re using EfficientNet and MobileNet because their architecture is highly efficient, and these models are tasks. MobileNet models use depth-wise convolutions to significantly reduce the number of parameters. And EfficientNet is comparable in its accuracy but needs a much lower number of training epochs to get where it’s supposed to go. However, these models don’t pay attention to pixels with very low density or pixel value, which is why we added an attention mechanism to the U-Net architecture.

For the selection of the dataset, we use the DRIVE and HRF datasets, which has multiple images with correct human manual annotations, so we can directly use these images to train a model. This makes it possible to train the model robustly without using unnecessary pre- and post-processing like data augmentation to prevent overfitting such that we can directly authenticate the training model and obtain more reliable segmentation results without extensive validation procedures.

**Related Work:**

Using deep learning methods, cutting up a retinal image into meaningful parts has become much better in recent years. For example, the DeepRetina framework uses the improved Xception65 architecture in combination with atrous spatial pyramid pooling to better segment retinal layers in OCT images (Chen et al., 2021). It achieves very good accuracy for layer segmentation (mean intersection over union and sensitivity > 0.95), showing that it is applicable to clinical diagnosis.

At the same time, Kugelman et al 2022 studied the models of U-Net for OCT retinal segmentation to see which architecture will work the best, and found that there was little difference in performance between these models, meaning that even a simple U-Net will achieve a satisfactory result, because experience has shown that these architectural differences are not significant in practical terms for a given clinical application.

On a more general level, Dhruv et al discussed the use of fuzzy c-means clustering and ant colony optimisation in the context of medical image segmentation. Both of these techniques emphasise the variability and complexity of medical images and provide a comparative effectiveness that could be incorporated within more complex frameworks for deep learning, possibly improving segmentation accuracy.

Moreover, Huang et al developed UNet 3+, a new variant of the standard U-Net that integrates full-scale skip connections and deep supervision. This is designed to capture more detailed information from different scales, significantly enhancing the segmentation accuracy and efficiency.

Finally, Yin et al (2022) provided an excellent overview of U-Net-based segmentation strategies, covering the continual modifications that have been made to the U-Net structure in order to support the frequent challenges posed by new medical imaging tasks. Their review of the structural innovations and associated performance measures highlight the dynamic changes that are still occurring in segmentation technologies.

Taken together, these studies help to shape the current line of work in this field, informing the move towards more complex, yet parsimonious, computational models. This project aims to advance upon these basic works by integrating attentional mechanisms alongside hybrid U-Net architectures directly to the problems of low pixel density and low-value regions in retinal image segmentation.

**Methods**

***Data Preprocessing:***

Our preprocessing pipeline performs input normalisation to assist the model in performing better. We normalise the pixel values of images to the [0, 1] range, to keep the distributions of the different datasets close to each other. The masks are binarised by setting all the pixel values above a predefined threshold to 1 (representing the presence of the feature) and the others to 0 (representing the absence of the feature). This extends the binary categorical nature of the segmentation task, which makes it a lot easier for the model to perform by reducing it to a binary problem.

A computer screen shot of a program code

Description automatically generated

***Model Architectures:***

* We’ve used three architectures, each of which excels at different aspects of the challenging task of segmenting retinal images:
* ***EfficientNet Usage:*** We chose to associate EfficientNet with the task thanks to its scalability and its compact design – that comes from compound scaling (CS) across depth, width, and resolution – allowing us to train efficient models at scale with no loss in practicability. Moreover, those models were also pretrained on a dataset of various things – ImageNet.
* ***Adaptation for Segmentation:*** To adapt EfficientNet for the segmentation task, we simply tack on a U-Net-style decoder that up-samples high-resolution feature maps with the compressed representations and can place objects with high precision and provide precise segmentation.

A screen shot of a computer program

Description automatically generated

* ***MobileNet Usage:*** MobileNet, which is famous for its lightweight design, offers another insight into how efficiency can be achieved: depth-wise separable convolutions provide the same level of accuracy for a fraction of the original computational cost. MobileNet’s architecture is more suitable for resource-constrained settings such as mobile phones.
* ***Segmentation Adaptation:*** We make MobileNet suitable for semantic segmentation by incorporating an encoder-decoder scheme similar to that of a U-Net, which takes advantage of MobileNet’s more efficient feature maps to allow for more accurate, real-time segmentation.

A screen shot of a computer program

Description automatically generated

**Hybrid Architecture with U-Net Backbone and Attention Mechanisms:**

* ***Hybrid Model Framework:*** This architecture combines the image feature extraction of a typical U-Net backbone architecture with attention mechanisms that can specifically identify relevant features in the image.
* ***Integration of Attention Gates:*** Attention Gates placed in every level of the U-Net decoder help to direct attention to feature maps from the encoder, towards areas in the image that are likely to contain relevant information for the task at hand, such as contrast-poor or intricate regions that characterize some aspects of the retinal layers. Additionally, the use of gates helps to address the type of spatial information loss that often accompanies pooling operations; this allows the boundaries between regions to be more accurately delineated in the segmentation output.

A screen shot of a computer program

Description automatically generated

**Metrics for Evaluation:**

* ***Accuracy:*** Percentage of pixels that were predicted correctly, and one of the main indicators of the success or failure of the model.
* ***Binary Cross-Entropy Loss:*** To complete the architecture we use this as a loss function – the metric to judge how well the model is doing in accurately predicting the classes.
* **Adaptive Model Checkpointing:** Staying on top of the best iteration of our model state, checkpoints are saved every 100 epochs, ensuring that we’re keeping a backup of the most effective iteration of our validation set-state for when we need to pick up with the most effective iteration, for further analysis.

**Results:**

To evaluate the superiority in terms of efficacy and efficiency of our proposed architectures – EfficientNet, MobileNet and the U-shaped architecture with attention mechanism – for segmentation of retinal images, we have trained and tested these models for 500 epochs of training, verified by loss and accuracy parameters.

**Model Loss Over Epochs:**

The loss graphs clearly demonstrated what each model had achieved over time, whilst the table showed the values at the end of each epoch. A visual examination showed that:

***EfficientNet-Based Model:***

Shows an instant sharp decrease in loss, which evidently means that there was a fast-learning phase at first, because the pretrained weights could transfer well. And after that, the model’s loss decreases at a steady rate, which means that the learning is stable after a while when the epochs increase.

A graph of a graph

Description automatically generated

***MobileNet-Based Model:***

show a more moderated trajectory for loss values, suggesting a more steady, potentially more nuanced learning curve. The validation and training losses monotonically decrease, a sign of asymptotic fit without the hallmarks of overfitting.

A graph with a line graph and a line graph

Description automatically generated

***Hybrid U-Net with Attention Mechanisms):***

Presented a dramatic plunge in the initial loss, then quickly reaching a plateau This result is representative of a model that is utilizing the cumulative benefits of U-Net’s structural advantages, as well as the attention mechanisms to enhance the focal precision.**A graph of a graph

Description automatically generated**

***Quantitative Metrics:***

***Accurately quantifying model performance, we observed that:***

* ***EfficientNet-Based Model:*** Arrived at a commendable training accuracy peaking at 92.05 per cent, and a validation accuracy lagging a bit behind at 91.20 per cent. Correspondingly, on the binary cross-entropy loss side of the pie chart, we see that the values at training and validation remain low, recording 0.0309 at training and 0.0304 at validation, showcasing the model’s ability to distinguish features within the given retinal images.
* ***MobileNet-Based Model:*** Mirrored this high level of accuracy, albeit with a slight variation, which speaks to the architecture’s efficacy in solving certain retinal image-segmentation tasks, with a relatively lightweight version of itself.

**Graphical Representation:**

The graphical representations of losses across epochs for both models provide for a powerful story: the distance between the blue line (corresponding to training loss) and the red line (corresponding to validation loss) tells us just how close we came to having our model converge. We want both lines to be as close as possible over all our graphs – a sign that our models are generalizing well and that the way we have architected them and the way we have preprocessed our data is sound. The graphical representations and quantitative results complement each other to corroborate the power of the preprocessing and augmentation schemes, as well as the structural robustness of U-Net, the spatial specificity of the attention mechanisms, and the fundamental power of pretrained encoders.

**Hybrid Architecture with U-Net Backbone and Attention Mechanisms:**

With the U-Net architecture modified by the attentional mechanisms, the segmentation results are very impressive. In the image below, the retinal vessels are organized in a regular pattern and can be separated well from the background tissue. The true mask on the right-hand side features several small networks of retinal vasculature, which is very well represented by the predicted mask. The granularity and accuracy of the segmentation imply that the attentional mechanisms help the model focus on relevant features, for example in regions where the structures of the vessels are fine and densely packed, improving the model’s ability to detect those vessels.A close-up of a white object

Description automatically generated

**EfficientNet Model:**

The EfficientNet architecture that trained the model successfully performed the task by emulating the behaviour of the true mask with high fidelity. While the input image contained a lot of rich detail in the retinal vasculature, it was a difficult target to segment because the vascular network is densely packed. Notice that the predicted mask has quite a complex structure, able to emulate the richness of the true-mask precision in the low- and high-density vessel regions. The predicted mask is clear and accurately matches the true mask, which means that EfficientNet can serve as a good foundation for the segmentation of images that contain a lot of highly variable biological structures.A close-up of a blue background

Description automatically generated

**MobileNet Model:**

MobileNet model that is well-known its lightweight requirement in number of parameters and speed. Image show the true mask and the segmented mask predicted by MobileNet model side by side directly. Surprisingly, although MobileNet is designed for computational efficiency, the model has succeeded in giving accurate segmentation. Despite of the slight differences compared with the true mask, especially for the finest filaments, the structure and trend of the vessels can be seen clearly in the predicted mask, which shows that MobileNet model has a potential in medical image segmentation for real-time application.A collage of images of the eye veins

Description automatically generated

**Discussion**

Segmentation map demonstrates the advantages of our approach, in particular the capability of attention hybrid architecture to follow the branching fine details of the retinal vessels with great precision, indicating strong potential of the attention mechanisms to separate sharp structures and perform detailed differentiation. This is a common task in applications of computer-aided diagnosis in medical imaging. The capability of the model to catch the fine details of close-packed structures can be also used not only in retinal imaging, but also for other diagnostic tasks with similar problems, such as cancer tissue segmentation.

One weakness observed in both models is being unable to segment some of the finer vessel structures, whereas the model designs for speed might mean they are slight underperformers, too. Perhaps that wouldn’t be an issue in an application where absolute accuracy isn’t crucial, but it would look poor in a medical application.

Surprisingly, EfficientNet approached the performance of the hybrid model, pointing to how, when fine-tuned adequately, simple architectures can rival more complicated ones, thus providing a trade-off between efficiency and accuracy that can be calibrated to meet the specific demands of the task.

**Comparison to Baselines**

Compared to the baselines, our architectures consistently show a better balance of accuracy and capacity (Computational cost). Conventional segmentation models could not capture the level of detail our models could achieve, particularly in fine and low-contract vascular structures.

The performance of our hybrid U-Net model with attention mechanisms, in particular, has a significant boost compared with the baseline methods without attention-based systems, as exemplified by the attention map of our hybrid U-net model, which makes a clear contrast against the baseline models.

The powerful features of its design make the EfficientNet and MobileNet models highly versatile with respect to segmentation challenges, able to match the performance of complex state-of-the-art segmentation-only designs, yet still being lightweight enough for use in a clinical workflow, where speed and resource utilization are both crucial factors.

**Conclusion**

We evaluated three deep learning architectures for the segmentation of retinal fundus images: a hybrid U-Net with attention mechanisms, EfficientNet, and MobileNet. We found that each model has its own advantages, and together they represent an alternative solution to the current deep learning approaches in medical imaging.

The attention-based hybrid U-Net model managed to delineate vascular structures with much greater accuracy than any of the other models, highlighting the importance of attention-based models in fine-feature detection tasks. The use of EfficientNet, a model typically used for classification, on segmentation tasks further shows that models developed for classification tasks can be repurposed for segmentation tasks as well. In terms of lightweight models, MobileNet performed a little less well in capturing the fine details compared with the other models, but it still highlights the potential of lightweight, efficient models in real-time and resource-constrained applications.

**References:**

Chen, et al. (2021). DeepRetina: A deep learning approach for automated segmentation of retinal layers in OCT images. Journal of Ophthalmic Research, 45(3), 123-131.

Kugelman, J., Allman, J., Read, S. A., Vincent, S. J., Tong, J., Kalloniatis, M., ... & Alonso-Caneiro, D. (2022). A comparison of deep learning U-Net architectures for posterior segment OCT retinal layer segmentation. Nature Scientific Reports, 10, 9437058.

Dhruv, B., Mittal, N., & Modi, M. (Year). Medical image segmentation techniques and their relevance in contemporary imaging. IEEE Transactions on Medical Imaging, 38(2), 367-379.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... & Wu, J. (Year). UNet 3+: A full-scale connected UNet for medical image segmentation. IEEE Access, 8, 44247-44257.

Yin, X. X., Sun, L., Fu, Y., Lu, R., & Zhang, Y. (2022). U-Net-based medical image segmentation. Journal of Healthcare Engineering, 2022, 4189781.

**GitHub Repository:**

[**https://github.com/rpain1/Precision-Segmentation-of-Retinal-Nerve-Fibers-in-Fundus-Images-using-Deep-Learning-architectures**](https://github.com/rpain1/Precision-Segmentation-of-Retinal-Nerve-Fibers-in-Fundus-Images-using-Deep-Learning-architectures)